Telling Multifaceted Stories with Humanities Data: Visualizing Book of Hours Manuscripts

Rongqian Ma[1] and Kai Li[2]

[1] University of Pittsburgh, Pittsburgh PA, USA
rom77@pitt.edu
[2] Drexel University, Philadelphia PA, USA
kl696@drexel.edu

Abstract. The Book of Hours is a type of book artifact that was produced and circulated in Europe during the Middle Ages and the Renaissance. Its name is from the fact that each book contains a sequence of prayers that were recited hourly throughout the day. This study examines the visualization of a dataset of Books of Hours collection, which was curated and digitized at the University of Pennsylvania Libraries. We performed a data cleaning and manipulation procedure on this manuscript dataset and created information visualization using the R programming language. Aiming to communicate to the wider audience the rich history of Books of Hours with meaningful visualizations, we focused on the temporality, spatiality, and decoration aspects of the resource. This project serves as a point of departure to engage in broader topics at the intersection of information visualization, digital humanities, and manuscript studies, especially how to leverage humanistic interpretations of data into visualization practices.

Keywords: Information Visualization, Digital Humanities, Book of Hours, Manuscript Studies, Archives.

1 Introduction

The Book of Hours is a type of manuscript that were "commissioned and produced, bought and sold, bequeathed and inherited, printed and reprinted than any other text" from the mid-thirteenth to the mid-sixteenth century, which "constitute[d] one of the most significant groups of cultural artifacts from the late Middle Ages and Renaissance" (Wieck, 1997). It is a book artifact that contains a sequence of prayers to Virgin Mary, the Mother of God, that "were recited throughout the course of the entire day, sanctifying it through her to God, Hour by Hour" (Wieck, 1997, p.9). Within the multiple centuries of evolution, the Book of Hours straddled the revolution of manuscript to print and its physical format varied across the time and geography. These features were wonderfully embodied in a digitized collection of Books of Hours at the University of Pennsylvania Libraries, collected and curated by scholars at the Schoenberg Institute for Manuscript Studies as part of the *Bibliotheca Philadelphiensis* project (University of Pennsylvania Libraries, 2019).

Utilizing this collection as the primary source, this study aims to explore how the data visualization methods and graphical representations of the manuscript data contribute to (1) displaying features of the digitized archival collection and (2) constructing multifaceted stories of Books of Hours. This case study also serves as a point of departure for future engagement with broader issues at the intersection of information visualization, digital humanities (DH), and digital manuscript studies, particularly the relationships between the design of graphs and the characteristics of data.

2 Literature Review

Data visualization has been a trendy topic in digital humanities. It has been instrumental in humanities scholarship, particularly in terms of raising new research inquiries, providing new forms of analysis, and revealing new discoveries and supporting arguments (Estill, Jakacki, & Ullyot, 2016). The abundance of case studies in DH also illustrates that data visualization engages in a broad range of DH topics and methodologies such as text analysis, topic modeling, spatial, network, and image analysis, assisting scholars in obtaining insights and addressing research questions. From a theoretical perspective, digital humanists discussed how to engage with visualization practices more effectively. Johanna Drucker (2011) discussed the problematics and principles of the humanities data visualization and argued for a treatment of humanities data as "capta" that are "*taken* not *given*, constructed as an interpretation of the phenomenal world, not inherent in it." Data are not "self-evident, value neutral, and observer-independent" and therefore should be conceived as "qualitative, co-dependently constituted — in other words, [as] *capta*" (Drucker, 2011).

Extending Drucker's (2011) assertation, Sinclair, Ruecker, and Radzikowska (2013) argued that "[a] humanities approach [to visualization] consists not of converging toward a single interpretation that information visualization for humanities scholars needs to accommodate a mix of evidence and cannot be challenged but rather of examining the objects of study from as many reasonable and original perspectives as possible to develop convincing interpretation." Therefore, the visualizations should be evaluated by "determining how well they support this interpretive activity." Here brings out the question about what is a good visualization for a humanities topic and how to leverage humanistic concerns such as complexity, sophistication, interpretation, and context (Borgman, 2007) into visual forms.

In the field of manuscript studies, digital methods and tools have also been widely utilized in research during the past decade (Estill, Jakacki, & Ullyot, 2016; Carson & Kirwan, 2014; Nelson & Terras, 2012). Several examples of the manuscripts visualization, ranging from the case of visualization of Arab manuscripts (El Bannay et al., 2009), the examination of the collation visualization system (Porter, Campagnolo, & Connelly, 2017), to the development of a visualization technique for high-dimensional medieval manuscripts (Chandna, Tonne, & Stotzka, 2016), demonstrate the role of information visualization as an important component of digital analyses in manuscript studies. Situated within the context of digital manuscripts, this study further explores

how data visualization assists storytelling of manuscripts and discusses what considerations should be taken into account for visualizing manuscript data.

3 Data and Methods

The dataset of this project consists of 185 entries containing descriptive information about three categories of metadata information about the Books of Hours manuscript collection: the dates of production, geographies of production, and decorations. The geographical scope of the data is western Europe while the temporal scope is 13th-19th centuries. The data categories and their content were obtained from the finding aid of the digitized collection, which were essentially selected and curated by researchers and curators of the Books of Hours manuscripts at Penn Libraries. Due to the fact that all the information was extracted from the original idiosyncratic manuscripts, data format varies tremendously across the dataset.

Facing the messiness of the dataset, we performed substantial data preparation tasks for the three categories of information before using software R, specifically the ggplot2 package, for data visualization. Major strategies include (1) data selection, (2) data merge, (3) data standardization and reformatting, and (4) text analysis. Every decision on data preparation was a reflection of the critical judgment of data and demonstrated researchers' interpretations of the dataset. A detailed account of the dataset characteristics and preparation strategies is presented in Table 1.

For the decorations specifically, we performed exploratory text analysis of the original data using the Voyant Tools (Welsh, 2014), so as to identify potential keywords for the visualization of the decorations. This was because the decoration information in this dataset is presented as textual descriptions. The text analysis of the decoration data generated two interesting sets of decoration, one that focused on decorative patterns such as illuminated initials and foliate borders, and the other on miniature themes such as the portraiture of Virgin Mary and Christ. Based on the text analysis results, a review of literature was further conducted to validate the selected keywords and ensure that they represented meaningful decoration traits of Books of Hours.

Table 1. Dataset summary and data preparation strategies.

Data Category	Feature of Category	Data Preparation Strategy	Example
Date	Single year, multiple years, time period, multiple periods, centuries.	• Merged single years into periods; • Separated multiple periods and merged them into corresponding single periods and centuries.	e.g., One record shows its date information as "late 15th century, with 18th -century additions" and we mapped it onto both 15th and 18th centuries
Geography	Regions (e.g., Eastern Mediterranean), countries (e.g., France), and cities (e.g., Paris); old and new European countries (e.g., Flanders, Bruges) coexist in the dataset.	• Merged all the cities into geographic regions/countries; • Old and new geographies coexist in the visualizations, demonstrating a mixture of temporality scales for Books of Hours.	e.g., Merged Paris into France; e.g., old geographies such as Flanders remained in the visualization and coexisted with more recent geographies
Decoration	Textual descriptions of decoration patterns and themes of miniature illustrations.	• Treated all decoration information as texts and performed keyword analysis using Voyant Tools, an open-source web-based application for text analysis; • Selected representative themes of miniature illustrations and decoration patterns from textual descriptions with Voyant analysis; • Coded all themes and decoration patterns for visualization in R programming language.	e.g., Frequent keywords such as "illuminated miniature," "illuminated initials," and "foliate" were captured by Voyant Tools and selected for visualization

4 Graphs and Findings

This section presents the visualizations created for the project and explores how these graphs assist in the communication of characteristics and features of Books of Hours in this collection. Aiming to explore the relationships between different categories of the dataset which cannot be necessarily revealed from browsing of the digitized collection and the raw dataset, we utilized heatmap graph to portray various stories of Books of Hours at the intersection of time, geography, and decoration categories. Using color-

coding to represent different values of the data, heatmaps are able to display the relationships more easily and construct inspiring and multifaceted historical stories about the Books of Hours. As demonstrated in the following heatmaps, each graph tells a bit of the stories for Books of Hours, adding to the histories of them from a perspective of manuscript archives.

4.1 Date and Geography

Time and geography are two fundamental units of information in this collection of Books of Hours and more broadly, humanities scholarship. However, both concepts have more situated implications in this specific case. Drucker (2011) argued that in order to recognize that all *data* in humanities are *capta* and to apply humanistic principles to data visualization, a specific approach is to incorporate the "relational structure" of humanities concepts into visualization, which is "situated in terms of different humanistic factors…and cannot always be treated equally and similarly" (Drucker, 2011). The notions of "time as temporality" and "space as spatiality" (Drucker, 2011) are such examples where the metrics of them are more sophisticated, subjective, multidimensional, and constituted.

Figure 1 and Figure 2 illustrate the production of medieval Books of Hours across area and time. The current visualizations shown below are the simplified version of date-geography intersectional heatmaps, presenting a mixture of historical and contemporary geographies. This attempt is to offer a relatively comprehensive (while simplified) landscape of the Books of Hours productions. The real relationships between geographies and times can be more complex with the temporal and geographical evolution of Books of Hours productions in Europe, but the synthesized presentations below offers an overview reading of the productions from the contemporary perspective, providing a basis for further explorations and more detailed visualizations of the collection data.

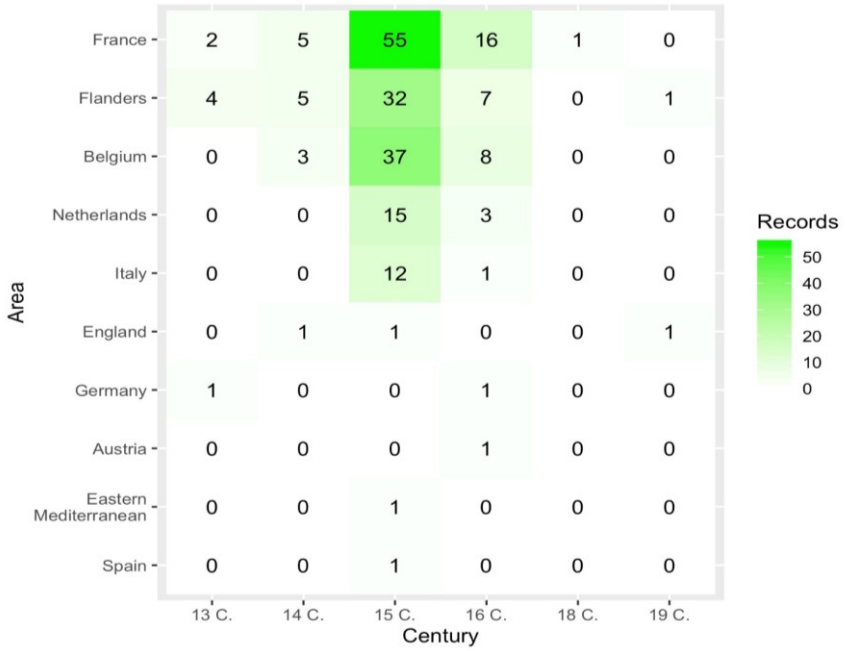

Fig. 1. Number of Records by Area and Century.

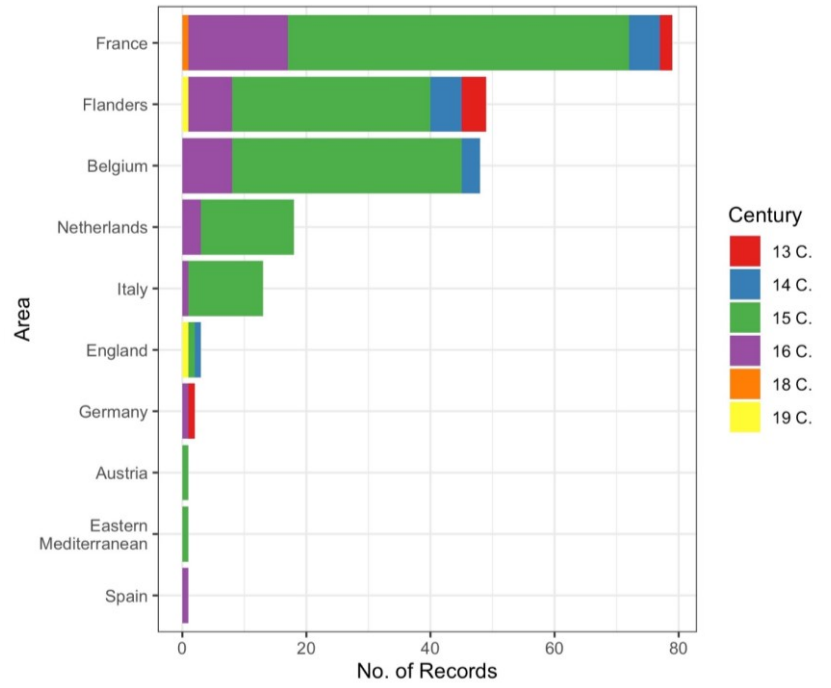

Fig. 2. Which area(s) created the most Books of Hours?

4.2 Decoration

Figure 3 shows how the decoration categories are used in the collection from 14th to 16th centuries. The color represents the ratio of records in each century using specific decorations. It reveals that border decorations, illuminated initials, and miniatures were the most popular and widely adopted decorative patterns during the 15th and 16th centuries. Figure 4 further elaborates on the use of decorations across geography in the collection. For the five areas that contribute the largest number of records to the collection, France and Italy utilized the most diversified types of decorative patterns while Books of Hours in Belgium seemed to be the least decorative compared with the others.

Another interesting group of information for this collection is the thematic decoration of miniatures, which may potentially demonstrate the receptions of religious images and portraiture in Books of Hours during the Middle Ages. Figure 5 and Figure 6 respectively demonstrate the use of miniature themes across time and area. The graphs illustrate that Annunciation and the image of Virgin Mary were the most frequently used themes of miniature during the 15th century while the portraiture of Saints was the most popular miniature theme in the 16th century. Compared with the temporal distribution of the themes, the use of miniature themes across geography tends to be more scattered. The Netherlands had a special preference in displaying the image of Christ in Books of Hours while Belgium – although valuing a relatively plain style with few decorative patterns, tended to emphasize the use of miniatures of various themes.

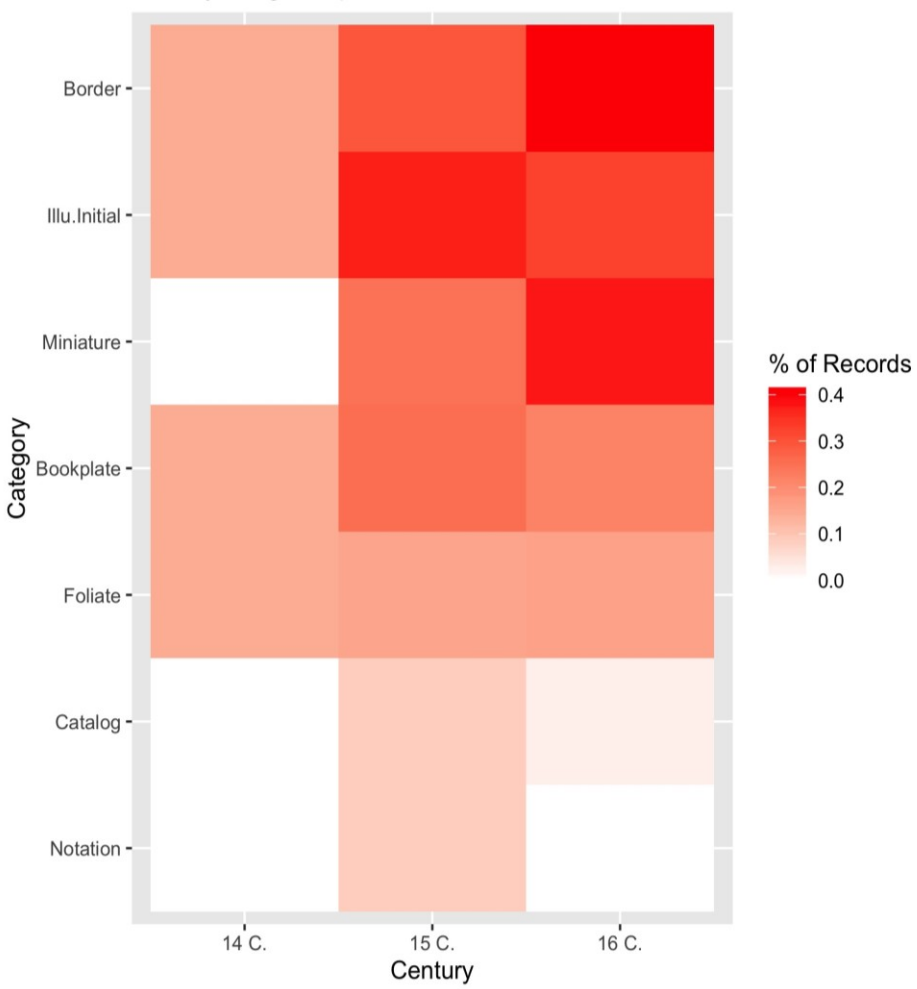

Fig. 3. Ratio of Records with Decoration by Century.

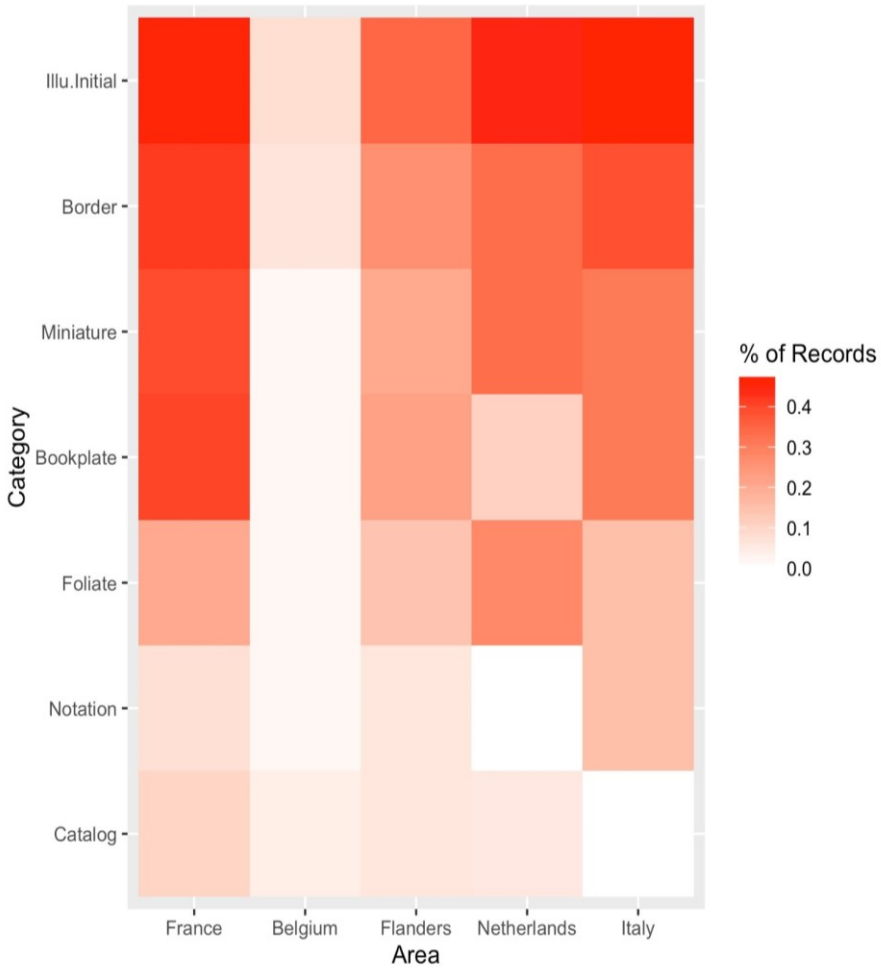

Fig. 4. Ratio of Records with Decoration by Area.

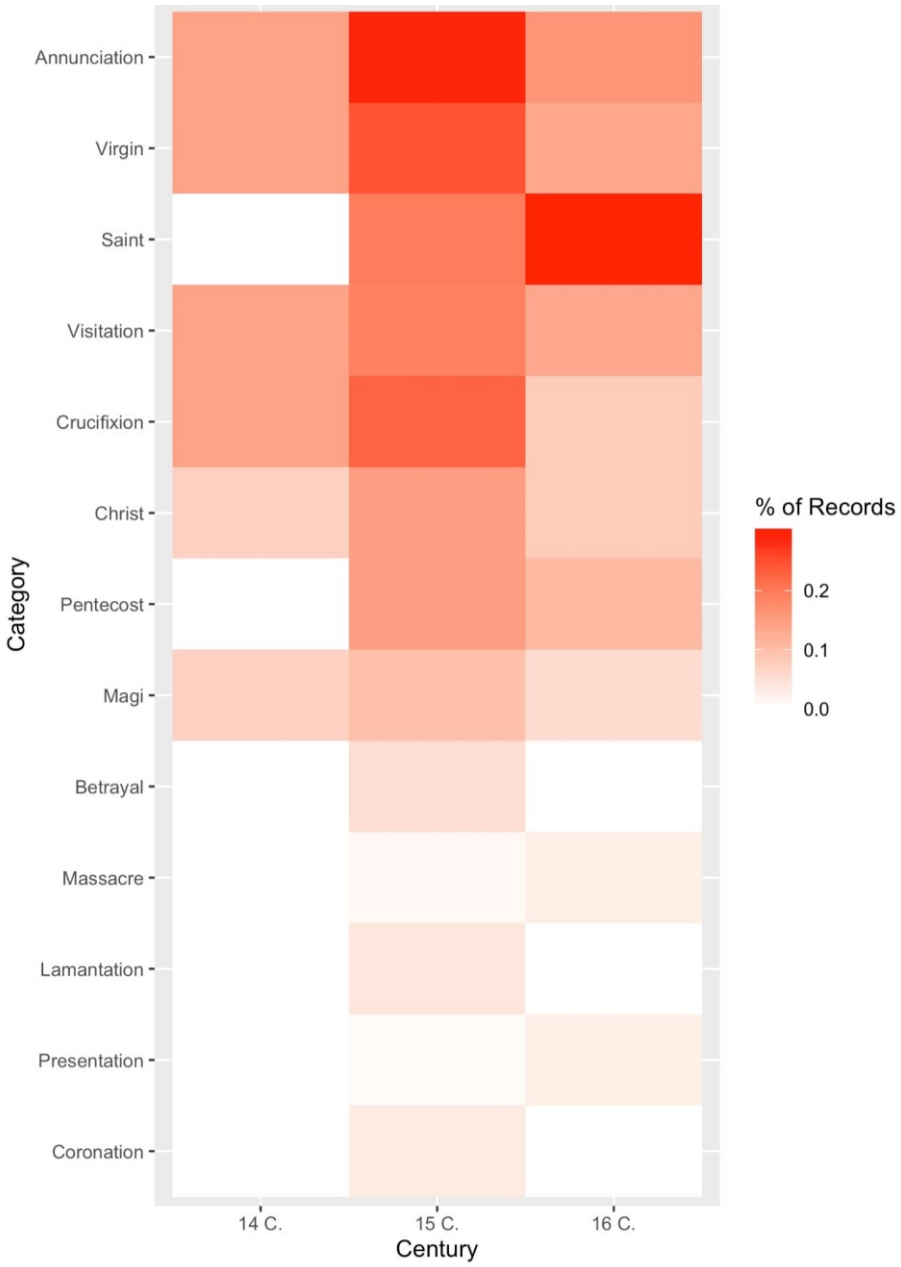

Fig. 5. Ratio of Records with Decoration Theme by Century.

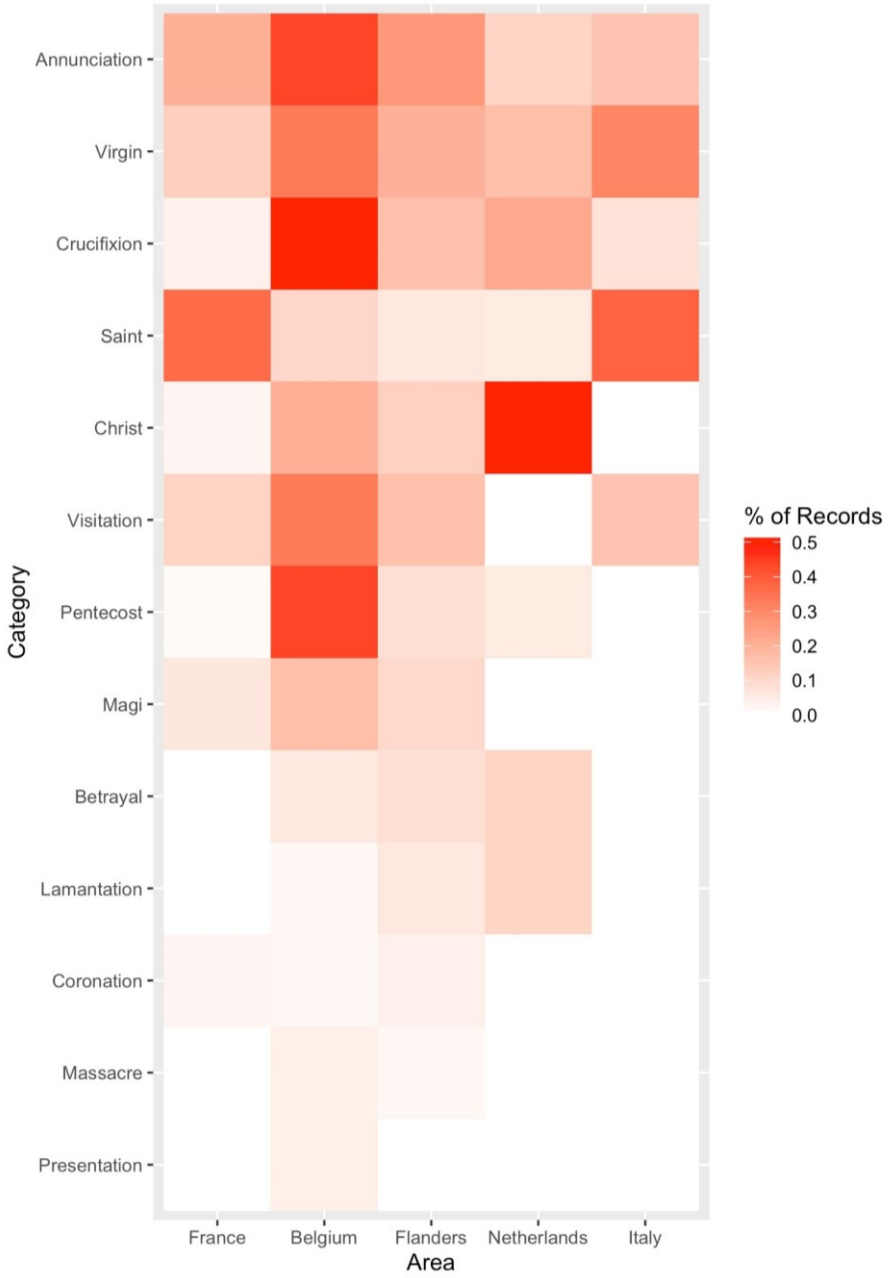

Fig. 6 Ratio of Records with Decoration Theme by Area.

From these graphs we see the characteristics of Books of Hours revealed from different aspects – particularly at the intersection of spatiality, temporality, and decoration. The visual representations of the manuscripts supplement the collection metadata as well as

the online finding aid as another way of introducing the archival collection and communicating the values of it to wider audiences. Besides, for researchers in humanities and manuscript studies, the visualizations may also help them explore information that is otherwise easily neglectable and examine the Books of Hours collection from various perspectives.

5 Discussion and Future Work

In this paper, we presented the visualizations created for the Books of Hours collection and discussed our observations about and solutions to the unique challenges of visualizing humanities data. Some of our observations are discussed in this section.

Rather than neutral techniques, information visualization has its own epistemological stances (D'Ignazio & Klein, 2016) that are not necessarily consistent with how humanists perceive the world. To emphasize their differences, Drucker (2011) discussed the significance of interpretation for humanities data, while Borgman (2007) further demonstrated that contexts, longevity, and diversity and complexity are distinctive features of humanities data. Anything that contains humanistic meaning in it can be treated as data in humanities scholarship (Borgman, 2007); and therefore, data practices such as visualization tend to be more individualistic and demand case-by-case considerations and decisions. Humanistic interpretations of data, as widely discussed in DH scholarship, is an important factor for successful data practices in humanities-related studies.

Data cleaning and preparation are significant components in this project. While proper data cleaning and wrangling assisted in creating clear accounts of stories for Books of Hours, we were constantly concerned about the possible harm such practices might bring to the study. For example, whether the complexities and variances of data are properly preserved and delivered in the visualizations was one key issue. Another consideration was to incorporate humanistic data interpretations into the visualization practice. The selection, reorganization, and visualization of certain bits of information for Books of Hours all demonstrate the critical judgment of the researchers on meaningful historical aspects of Books of Hours. How to effectively leverage data interpretations into established data visualization principles and standards is one issue faced by digital humanists and one that demands further research.

Although the current study demonstrates great potentials of using visualization methods to examine and communicate Books of Hours manuscript data, more issues can be further addressed to tell better stories of this book artifact. First, more information can be mined from the dataset, such as the length of the manuscripts, sequence of prayers, ownership of the manuscripts, and relationships between each manuscript. Such information is suggested in the dataset in various ways (e.g., titles of prayers and their page location, and book stamps) and has the potential to delineate other interesting stories of Books of Hours as well as furthering research on this topic. Second, extending from the single case study, a more comprehensive and in-depth discussion of information visualization principles for data related to humanities subjects will contribute theoretically to effective visualization design for humanities data, or the manuscript

data in particular. Finally, another direction of research is to further engage the humanists to validate our approaches and understand what they need from visualization methods.

References

1. Borgman, C.: Scholarship in the digital age: Information, infrastructure, and the Internet. The MIT Press, Cambridge (2007).
2. Carson, C., Kirwan, P. (eds.): Shakespeare and the digital world. Cambridge University Press, Cambridge, UK (2014).
3. Chandna, S., Tonne, D., Stotzka, R.: An effective visualization technique for determining co-relations in high-dimensional medieval manuscripts data. Electronic Imaging 2016(1), 1-6 (2016).
4. D'Ignazio, C., Klein, L. F.: Feminist data visualization. In Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore. IEEE (2016).
5. Drucker, J.: Humanities approaches to graphical display. Digital Humanities Quarterly 5(1) (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html
6. El Bannay, O., Rais, N.: Application for visualization of Arab manuscripts. In 2009 Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, IEEE (2009).
7. Estill, L., Jakacki, D., Ullyot, M.: Early modern studies after the digital turn. Iter Press, Toronto, Canada (2016).
8. Nelson, B., Terras, M. (eds.): Digitizing medieval and early modern material culture. Iter Press, Toronto, Canada (2012).
9. Porter, D., Cam33pagnolo, A., Connelly, E.: VisColl: A new collation tool for manuscript studies. In Busch, H., Fischer, F., Sahle, P. (eds.) Codicology and Palaeography in the digital age, vol. 4, pp. 81-100. BoD, Norderstedt, Germany (2017).
10. R Core Team.: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2019). URL https://www.R-project.org/
11. Sinclair, S., Ruecker, S., Radzikowska, M.: Information visualization for humanities scholars. In Literary studies in the digital age: An evolving anthology. Modern Language Association of America (2013) https://dlsanthology.mla.hcommons.org/information-visualization-for-humanities-scholars/.
12. University of Pennsylvania Libraries *BiblioPhilly* homepage, http://bibliophilly.library.upenn.edu/?keyword=Book+of+Hours, last accessed 2019/09/22.
13. Voyant Tools homepage, https://voyant-tools.org/, last accessed 2019/09/22.
14. Welsh, M.: Review of Voyant Tools. Collaborative Librarianship 6(2), 96-97(2014).
15. Wickham, H.: ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York (2016).
16. Wieck, R.: Painted prayers: The book of hours in medieval and Renaissance art. George Braziller, Inc., New York (1997).